7

PädiUAQ

Revista de Proyectos y Textos Académicos en Didáctica de las Ciencias y la Ingeniería

UNIVERSIDAD AUTÓNOMA DE QUERÉTARO
FACULTAD DE INGENIERÍA

AÑO 4 NÚMERO 7

DIRECTORIO

PädiUAQ. Revista de Proyectos y Textos Académicos en Didáctica de las Ciencias y la Ingeniería.

Año 4. Núm. 007, junio de 2020, es una publicación semestral editada y publicada por la Universidad Autónoma de Querétaro, División de Investigación y Posgrado de la Facultad de Ingeniería.
Centro Universitario, Cerro de las Campanas s/n, Col. Las Campanas, C.P. 76010, Tel. (442) 192-12-00, ext. 7035.

Reserva de Derechos al Uso Exclusivo
No. 04-2017-040313301800-203
ISSN: En trámite

Ambos registros están en trámite por el Instituto Nacional de Derechos de Autor.

Las opiniones expresadas por los autores no necesariamente reflejan la postura del editor de la publicación.

PädiUAQ

Revista de Proyectos y Textos Académicos en Didáctica de las Ciencias y la Ingeniería

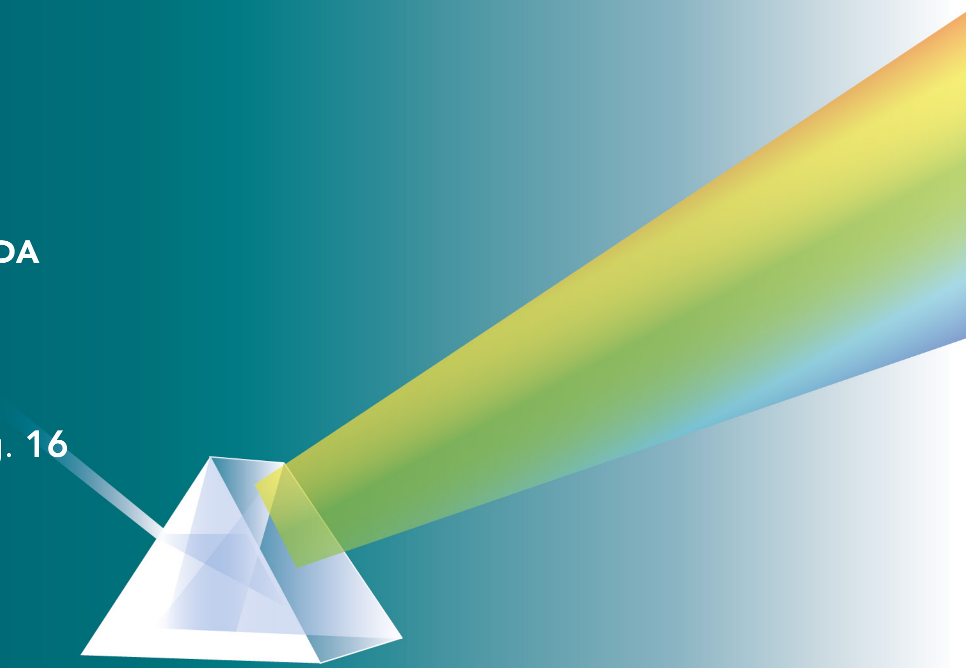
FACULTAD DE INGENIERÍA

COMITÉ EDITORIAL

PädiUAQ
Revista de Proyectos y Textos Académicos en Didáctica de las Ciencias y la Ingeniería

FACULTAD
DE INGENIERÍA

ÍNDICE



03

UN ANÁLISIS COMPARATIVO DE MODELAJE ESTADÍSTICO EN ESTUDIOS DE VIDA DE ANAQUEL SENSORIAL

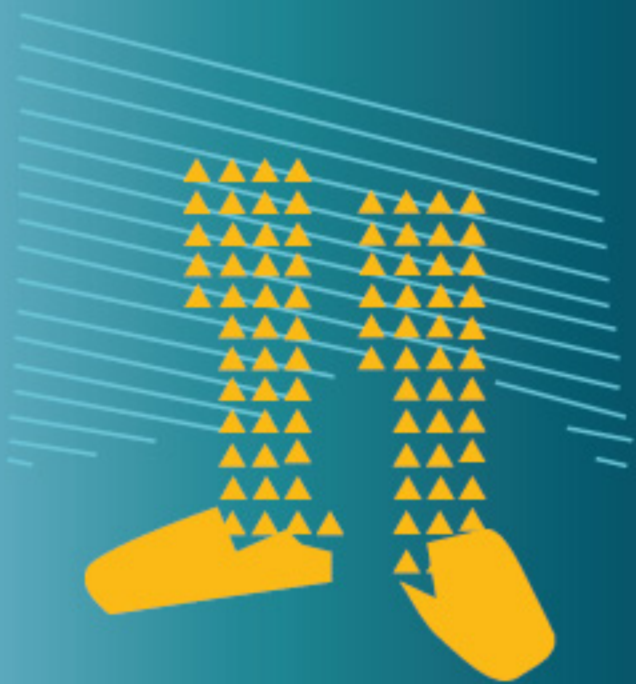
ILIANA MARÍA PATERNINA ORTEGA
EDUARDO CASTAÑO TOSTADO
MARIO SANTANA CIBRIÁN

pág. 24

MONTSERRAT LINO

UNIVERSIDAD AUTÓNOMA DE QUERÉTARO

LINOMONTSE@GMAIL.COM



01

COMPRENSIÓN DE LA PROBABILIDAD CONDICIONAL A TRAVÉS DE LA TEORÍA ANTROPOLÓGICA DE LO DIDÁCTICO

RESUMEN

Parte de la problemática descrita en la literatura especializada en el proceso enseñanza-aprendizaje de la probabilidad condicional está relacionada principalmente con los sesgos y el contexto del problema. La investigación descrita en este artículo tiene un enfoque cualitativo y exploratorio bajo observación natural; la muestra será por conveniencia, se obtendrá de un aula de estudiantes universitarios de la facultad de ingeniería que cursen probabilidad y estadística como parte de las materias de tronco común y que hayan cursado el tema de probabilidad condicional. Se asignarán a los participantes cuatro tareas, cada una con cuatro consignas; dos de las tareas se seleccionaron de cuestionarios que ya están estadísticamente validados para encontrar sesgos asociados al concepto de probabilidad condicional y son de contextos con los que el estudiante está familiarizado; mientras que las tareas restantes tendrán como característica la presentación de los problemas en un contexto distinto a su campo de conocimiento, pero evidenciarán el mismo sesgo. Las tareas se modificaron bajo el marco teórico de la Teoría Antropológica de lo Didáctico con la intención de que, al analizar sus producciones cognitivas, se detecte qué y cómo resuelven los estudiantes las actividades propuestas, aportando evidencia que permita establecer cómo comprenden la probabilidad condicional. Con la información obtenida se espera que sea posible focalizar los puntos donde es necesario incidir para la corrección de sesgos, y de esta manera, puedan resolver con éxito problemas de probabilidad condicional.

Palabras Clave: 000, probabilidad condicional, TAD, ACODESA, sesgo temporal, sesgo de causalidad.

ABSTRACT

Part of the problem described in specialized literature about the teaching-learning process of conditional probability is mainly related to biases and the context of the problems. The research described in this article has a qualitative and exploratory approach under natural observation. The sample will be chosen by convenience and will be obtained from a group of engineering students, at university level, who study probability and statistics as part of the common core subjects and who have studied conditional probability. Participants will be assigned four tasks, each with four items to fullfill; two of the tasks were selected from questionnaires that are already statistically validated to find biases associated with the concept of conditional probability and with which students are already familiar; while, in the remaining tasks the problems will be presented in a different context from their field of knowledge but they will show the same bias. The tasks were modified under the theoretical framework through the Anthropological Theory of Didactics, with the intention that, when analyzing their productions, it is detected what and how the students solve the proposed activities, providing evidence that allows to establish how they understand the conditional probability. With the information obtained, it is expected that the students acquire the capacity to focus on the points where it is necessary to influence the correction of biases, and in this way, they will be able to successfully solve conditional probability problems.

Keywords: 000, conditional probability, ATD, ACODESA, temporary bias, causality bias.

INTRODUCCIÓN

La probabilidad condicional según Díaz & de la Fuente [1], Barraguéz & Guisasola [11] es un tema importante debido a que se utiliza con frecuencia en la vida profesional y en la vida cotidiana, principalmente cuando se busca tomar decisiones acertadas en situaciones de incertidumbre, y también porque forma parte de la definición de otros conceptos estadísticos que son requisito en el estudio de la inferencia estadística, tanto clásica como bayesiana; así como en el estudio de la asociación de variables, la regresión y el contraste de las pruebas de hipótesis, entre otros conceptos que se definen mediante una probabilidad condicional. Por tanto, es imprescindible que el alumno adquiera las competencias y conocimientos básicos en este tema antes de avanzar en el estudio de la estadística.

Mejía, Sierra y Fernández [2] sugieren que parte de la dificultad aparece incluso desde la redacción de los problemas, ya que entenderlos puede resultar complejo, lo que obstaculiza su resolución adecuada. Esto ocurre principalmente por el contexto al que los problemas hacen referencia y con el cual los estudiantes pueden no estar familiarizados. Se considera que, cuanto mayor sea el número de probabilidades condicionales que contiene la parte formativa de un problema, mayor será la dificultad de los estudiantes para resolverlo.

La investigación teórica y experimental que se hace actualmente en Didáctica de la Matemática surge de observar que el alumno se equivoca cuando se le pide realizar ciertas tareas; cabe destacar que estos errores no obedecen a un modo aleatorio e impredecible, sino que ya se han focalizado; por otra parte, se ha determinado que otras dificultades se deben, según Batanero, Godino, Green, Holmes y Vallecillos [3], a la falta del conocimiento básico necesario para la comprensión correcta de un concepto o procedimiento dado. Por su parte, Sánchez [4] refiere que los errores o dificultades de los estudiantes en la resolución de problemas con frecuencia se deben a que tienen sus propias ideas previas sobre las situaciones, las cuales suelen ser persistentes e inconsistentes con el punto de vista normativo.

Las principales dificultades para comprender el concepto de probabilidad condicional (sesgos en el razonamiento, Díaz & de la Fuente [1] y Tarr & Lannin [5]) son:

• Influencia de la presentación del problema.

- Estructura que presentan los problemas de enunciado verbal.

- Relación con el contexto (no le es familiar al estudiante).

• Condicionamiento y causación.

• Condicionamiento y temporalidad (intercambio de sucesos en la probabilidad condicional): $P(A|B)$ es diferente de $P(B|A)$; este error es muy común cuando se trata de los errores en la interpretación del nivel de significación y el p-valor.

• Confusión entre probabilidad condicional y conjunta, pues depende de cómo se redacten los enunciados.

• Situaciones Sincrónicas (situaciones estáticas, donde los experimentos aleatorios se realizan simultáneamente) o situaciones diacrónicas (hay una clara secuencia temporal).

• Concepto de independencia.

Díaz & de la Fuente [6] refieren que la probabilidad condicional fue reportada por docentes universitarios como un concepto que se les dificulta comprender a los estudiantes. Con base en esto surge la siguiente investigación, cuyo objetivo es dar respuesta a la pregunta de investigación: ¿Cómo comprenden la probabilidad condicional los estudiantes universitarios?; para ello, se utilizarán algunas tareas propuestas para medir sesgos de falacia temporal (situación diacrónica) y de concepto de causalidad en la resolución de dichos problemas, los cuales han sido validados estadísticamente y pueden ser usados por otras investigaciones con dicho fin (estos problemas serán los descontextualizados). La presente investigación sólo analizará dos de los sesgos antes mencionados, debido a que, al ser una investigación de tipo cualitativo, no es posible analizar las producciones de muchos alumnos por el trabajo inherente definido por la TAD, la cual guía y sustenta toda la investigación y será considerada en cada una de las cinco fases del método Acodesa para su aplicación.

Los problemas elegidos se modificaron a la luz del marco teórico de la Teoría Antropológica de lo Didáctico (TAD) de Chevallard [7], la cual funge como eje rector de toda la investigación con la finalidad de cubrir el objetivo principal de caracterizar la actividad matemática de los estudiantes universitarios al resolver problemas de probabilidad condicional, analizando el qué y cómo resuelven estos problemas y aportando una idea de cómo es que comprenden la probabilidad condicional, pregunta de investigación que se pretende responder.

La investigación está acotada únicamente a la probabilidad condicional excluyendo el teorema de Bayes. Se remite a la que se enseña en los cursos básicos de probabilidad, con escenarios discretos finitos y equiprobables.

Aunque es un problema ya estudiado, no se ha hecho desde la TAD y la metodología Acodesa para su aplicación, que nos pueden dar otro tipo de sugerencias para la enseñanza-aprendizaje del concepto de probabilidad condicional. La razón de que la propuesta no haya sido probada y, por tanto, no se presenten resultados se debe a que es una investigación en curso y será probada antes de finalizar este año con alumnos de la FI-UAQ, como se definió en la muestra. Por el momento es sólo una propuesta; sin embargo, aunque no se tienen aún resultados en las investigaciones, es posible escribir el artículo enfocándose en el diseño del instrumento y cómo éste permitirá recoger lo que se pretende observar, indicándolo en la sección de resultados.

Con la información obtenida, será posible focalizar los puntos donde es necesario incidir para la corrección de sesgos a través del rediseño de una secuencia de tareas de probabilidad condicional mediante una selección de éstas de acuerdo a su grado de dificultad, con la finalidad de favorecer la comprensión progresiva del concepto y, de esta manera, lograr que resuelvan con éxito problemas de probabilidad condicional en cualquier contexto mediante la comprensión de los conceptos asociados y no por la memorización de fórmulas.

MARCO TEÓRICO

La Teoría antropológica de lo didáctico

Como referente teórico se utiliza la Teoría Antropológica de lo Didáctico (TAD) de Chevallard [7], la cual es una herramienta para el análisis del conocimiento matemático visto como un conjunto de prácticas sociales institucionalizadas que requiere de una forma de análisis que permita la descripción y el estudio de las condiciones de la realización desde la Organización Matemática. La Organización Matemática (OM) o Praxeología está compuesta por varios tipos de tareas cuya realización requiere técnicas matemáticas, que a su vez se justifican en tecnologías y teorías matemáticas específicas.

Tabla 1. Elementos de la Teoría Antropológica de lo Didáctico

Praxeologías Para que el conocimiento matemático se desarrolle hay que crear una situación **m(π, τ, θ, Θ)**	**(s) Tipos de Situaciones** π **Problemas** (actividad que involucra al objeto matemático)	**Praxis o conocimiento** **SABER HACER**
	τ **Técnicas** *(procedimiento que permite resolver la tarea)*	**Maneras de HACER** *procedimientos que pueden ser empleados para resolver los problemas*
	θ **Tecnologías** *(discurso que justifica y explica la técnica)*	**SABER** Discursos que sustentan, describen, explican y justifican los procesos matemáticos que ahí se encuentran involucrados y los cuales se espera que sean institucionalizados en los procesos de enseñanza-aprendizaje.
	Θ **La teoría** *(discurso que justifica y explica la tecnología)*	**SABER** Argumento formal que permite justificar *rigurosamente* la tecnología

Fuente: Modificado de Chevallard [7].

En la Organización Matemática o Praxeología se observan las prácticas matemáticas; éstas son actuaciones particulares en el abordaje de problemas matemáticos específicos. Está determinada por formas de razonar, comunicar, validar o generalizar, y habitualmente no existe de manera aislada.

Lo que este artículo presenta es una construcción de una praxeología matemática que permita determinar que los objetos matemáticos surgen de prácticas con las matemáticas ubicadas en diversos contextos geográficos y culturales, relacionada con la probabilidad condicional y direccionada por la TAD exponiendo la tarea, la técnica, la tecnología y la teoría.

Con la TAD, se podrá determinar si resolvieron correctamente el problema mediante el uso de alguna de las técnicas (τ) fundamentadas en alguna tecnología (θ) y alguna teoría (Θ).

METODOLOGÍA

La presente investigación tiene un enfoque cualitativo y exploratorio bajo observación natural (León & Montero [8]); la muestra será por conveniencia, se obtendrá de un aula de estudiantes universitarios de la facultad de ingeniería que cursen probabilidad y estadística como parte de las materias de tronco común y que hayan cursado el tema de probabilidad condicional. El instrumento de recolección de datos estará conformado por 4 tareas de probabilidad condicional con 4 consignas cada una; dicho instrumento será aplicado bajo el método de ACODESA (Hitt, González-Martín [9]), por su acrónimo en francés. Es un método de enseñanza para introducir conceptos matemáticos en el aula y promover el desarrollo del pensamiento matemático que se refiere al aprendizaje colaborativo, debate científico y auto-reflexión; además, provoca que los estudiantes construyan significados e involucra escenarios individuales inmersos en los socioculturales. En esta metodología, cada tarea constará de 5 fases: 1) *Trabajo individual*, donde se recuperan los conocimientos previos del estudiante pues su producción cognitiva produce significados propios correctos o no; 2) *Trabajo en equipo*, se espera un refinamiento de sus respuestas sobre la misma tarea a partir de sus procesos de argumentación y validación; 3) *Debate*, todo el grupo discute las distintas formas en la que resolvieron la tarea, se espera obtener un refinamiento de su técnica y eventualmente su teoría; 4) *Auto-reflexión*, proceso individual de reconstrucción para afianzar lo logrado en el debate, y 5) *Institucionalización del concepto* probabilístico asociado a la tarea analizada tomando en cuenta las producciones de los estudiantes. Cada una de las fases será dirigida y analizada bajo el marco teórico de la TAD. Se estima una sesión de 50 minutos para cada tarea.

RESULTADOS Y DISCUSIÓN

Para ejemplificar la construcción de la praxeología matemática asociada a las tareas, sólo se desarrolla este proceso para el caso de la Tarea 1, que forma parte de la secuencia estructurada de tareas para el estudio de investigación. Se pondrá otra tarea que evidencie el mismo sesgo diacrónico que la tarea 1 pero en un contexto del ámbito ingenieril, y otras dos que evidencien el sesgo causal.

La tarea 1 que se aplicará tiene por objetivo ver si se presenta la falacia temporal o sesgo de secuencia temporal, dado que es una situación diacrónica que fue modificada de Ojeda [10] (Fig. 1). Y se explica a detalle su análisis bajo la TAD. En la consigna c), la intención de hablar de 100 bolas se puso con el fin de que, al imaginar más bolas, le fuera más fácil al estudiante identificar lo que ocurriría, por la consecuencia de los grandes números y el comportamiento estadístico que conlleva pensarlo usando más de solo 1 bola.

Tarea 1. Analiza el diagrama y resuelve lo que se te pide.

Una bola se suelta por E. Si sale por R, ¿cuál es la probabilidad de que haya pasado por el canal I?

a) Escribe cuáles son los eventos y cuál es la probabilidad pedida.

b) Realiza esquemas que te ayuden a calcular la probabilidad pedida. Justifica tu respuesta.

c) Si en lugar de una bola tienes 100, ¿cambia el resultado?

d) Explica tu resultado. ¿Por qué obtuviste ese resultado?, ¿cómo sabes que es correcto?

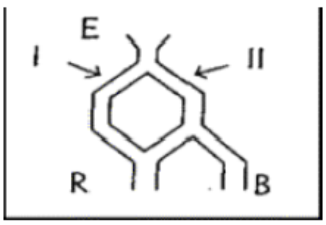


Figura 1. Actividad sobre sesgo temporal. Situación Diacrónica. Contexto no ingenieril. Modificada de Ojeda [10]

El análisis se hará con base en el marco teórico de la TAD.

Modelo praxeológico de Chevallard (T, τ, θ, Θ):

Tipo de Tarea (T). Calcular la probabilidad condicional.

Técnicas (τ). Aquí se describen a detalle las posibles respuestas que los estudiantes podrían dar a cada una de las consignas de la tarea (manera de resolver), donde todas las técnicas conllevan al mismo valor numérico:

τ 1a: Escribir los eventos:

I: Bola paso por el canal izquierdo (I)

R: Bola salió por R

P (I|R): Probabilidad Condicional de que la bola haya pasado por el canal I, dado que salió por R.

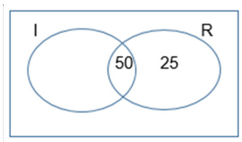
τ 2b: Realice el cálculo con la fórmula de probabilidad condicional $\boldsymbol{P(I|R)=P(I\cap R)/P(R)}$

Las técnicas 3 y 4 podrían ser referidas por estudiantes que han comprendido el concepto de probabilidad condicional P(A|B) como la probabilidad de que ocurra el evento A dado que ha ocurrido el evento B (que es el evento seguro), entonces puede entender que su espacio muestral queda reducido al evento B y ahora serán todos sus casos posibles; es decir, obsérvese que en A ∩ B ocurren conjuntamente A y B, y tiene asociado el espacio muestral original Ω. A|B significa que, en los casos en los que ya ha ocurrido B, ocurre A, y por tanto el espacio muestral es aquel en el que ha ocurrido el evento B.

τ 3b: Técnica de la regla de tres. 75 →100 y 50→$\boldsymbol{x}$ = (100*50)/75 o x = (5000/75)

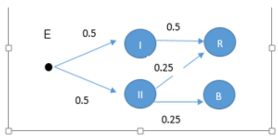
τ 4b: Dibuje un diagrama de VENN. Dejando R como espacio muestral, que es nuestro punto interés y no influye en el resultado el quitar B (ver Figura 2)

τ 5b: Dibuje un diagrama de árbol (ver Figura 3)



$$\boldsymbol{P(I|R)=\frac{P(I\cap R)}{P(R)}=\frac{50/_{100}}{75/_{100}}=\frac{1/_2}{3/_4}=\frac{2}{3}=0.66}$$

Figura 2. Diagrama de Venn y cálculo de la probabilidad condicional usándolo.



$$\boldsymbol{P(I|R)=\frac{P(I\cap R)}{P(R)}=\frac{0.5}{0.5+0.25}=\frac{0.5}{0.75}=\frac{2}{3}=0.66}$$

Figura 3. Diagrama de árbol y cálculo de la probabilidad condicional usándolo

Como refieren Lonjedo & Huerta [12], la presentación de los datos en algunos problemas de probabilidad condicional permite resolver estos problemas sólo con los requisitos del conocimiento numérico, tales como razón, proporción, etc., por lo que su estudio también implica el pensamiento numérico, pero la tradición en la enseñanza de las probabilidades nos muestra lo contrario. Los profesores no suelen utilizar el pensamiento numérico en el contexto de la probabilidad. Tal es el caso de la anterior técnica 3 y la técnica 6, que se muestra a continuación:

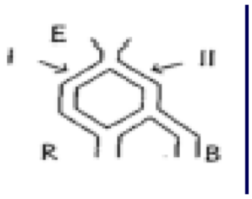
Tabla 2. Τ6b. Tabla de doble entrada. Renglones y columnas son independientes

Canal	R	B	Total
I (Izquierdo)	50 •	0 ♦	50
II (Derecho)	25 *	25 ▲	50
Total	75 §	25 ◘	100

Simbología: §= P(R); ◘= P(B) ; • P(R ∩ *I*); ♦ P(B ∩ *I*); *= P(R ∩ *II*); ▲= P(B ∩ *II*)

$$\boxed{\boldsymbol{P(I|R) = \frac{P(I\cap R)}{P(R)} = \frac{50/100}{75/100} = \frac{1/2}{3/4} = \frac{2}{3} = 0.66}}$$

τ $6b_1$: Usando la Tabla 2, podría haber usado: (•/§) = 5075 = 23 = 0.66
τ 7b: Sólo observa el esquema y apunta las cantidades, y sobre eso deduce:



$$\frac{50}{75} = \frac{2}{3} = 0.66$$

Figura 4. Esquema del problema

Tecnologías (θ). Tienen un carácter justificativo que argumenta la técnica realizada por el estudiante.

θ 1: Operaciones con conjuntos y operaciones de probabilidad
θ 2b: Uso de la fórmula de probabilidad condicional.
θ 3b: Noción de variables proporcionales (las cantidades que se corresponden se obtienen multiplicando una por un número constante; función f: = x → ax; la división como inversa de la multiplicación.
θ 4b: (θ1)
θ 5b: Técnica gráfica de diagrama de árbol.
θ 6b y θ $6b_1$: Haga uso de frecuencias observadas y esperadas.
θ 7b: No se observan tecnologías.

Teorías (Θ). Es un nivel superior de justificación de la tecnología usada y puede estar dada por teoremas, lemas, axiomas o definiciones puntuales (conceptos matemáticos).
Θ 1: Teoría sobre conjuntos y axiomas de probabilidad.
Definición de probabilidad simple, interacción de eventos.
Θ 2b: Definición de probabilidad condicional y uso de fórmula $(I|R)=\boldsymbol{P(I\cap R)/P(R)}$
Θ 3b: Teoría de los números racionales y sus operaciones.
Θ 4b: Teoría de probabilidades y conjuntos. Concepto de espacio muestral.
Θ 5b: Reglas de multiplicación para eventos no independientes.
Θ 6b y Θ $6b_1$: Concepto de probabilidades marginales.
Θ 7b. No se observan teorías.

Se pondrá otra tarea que mida el mismo sesgo de temporalidad diacrónica, pero en un contexto que le sea familiar al estudiante. Los resultados se podrán comparar con el antes descrito para contestar a nuestra hipótesis de que el alumno podrá identificar y resolver, mejor y sin sesgos, los problemas de probabilidad condicional en situaciones a las que está familiarizado que los que representan situaciones en otros contextos de uso. Lo mismo se hará con las otras 2 tareas restantes, que evidenciarán el sesgo de concepto de causalidad en dos contextos familiares y en uno no familiar. Como se describe en el apartado de metodología, la aplicación de las tareas se realizará según la lógica de la metodología ACODESA.

CONCLUSIONES

Se espera obtener resultados similares a otras investigaciones bajo el mismo problema, debido a que la compresión de la relación de condicionalidad se dificulta si la secuencia temporal de los sucesos no coincide con el orden dado en el condicionamiento. El suceso condicionante (caer en el orificio R) es posterior en el tiempo al suceso cuya probabilidad se evalúa (pasar por I). La secuencia temporal dificulta la identificación correcta del espacio muestral del experimento: {(I,R), (II,R0, (II,B)}, puesto que a R llega el doble de bolas desde I que desde II; P(I|R)=2/3 y los estudiantes dan erróneamente la respuesta 0.5, no teniendo en cuenta las bolas que caen en el orificio B (Díaz & de la Fuente [6], Barragués & Guisasola [11]).

Será importante ver si no darle al estudiante los resultados de posibles respuestas, como era la propuesta original de la tarea de donde fue tomada, y permitirle desarrollarla de una forma libre, provoca que no caiga en el sesgo que se está midiendo.

El efecto de la propuesta al usar la TAD y la metodología Acodesa para su aplicación evidenciará el tipo de técnicas y recursos didácticos, así como teorías y tecnologías que los alumnos suelen usar para resolver estos problemas, con la finalidad de homogeneizar su uso y enriquecer el concepto. El método Acodesa proporcionará una forma más sencilla de llegar a la comprensión e institucionalización del concepto de probabilidad condicional, al permitir a cada estudiante trabajar con otras personas (equipo) y grupalmente (debate), e incentivar el aprendizaje colaborativo; de esta forma, se observará la evolución de la comprensión del concepto de probabilidad condicional en cada una de las fases, lo que podría dar pauta a la mejor manera de enseñarlo.

Lo que evidencien los resultados permitirá actuar en la corrección de sesgos, rediseñando una secuencia de tareas de probabilidad condicional mediante una selección de éstas de acuerdo con su grado de dificultad, con la finalidad de favorecer la comprensión progresiva del concepto y, de esta manera, lograr que los problemas de probabilidad condicional se resuelvan con éxito en cualquier contexto.

AGRADECIMIENTOS

A la Dra. Lilia Patricia Aké Tec, asesora del trabajo de tesis que está en proceso, y a la Mtra. Luisa Granados Ramírez, co-asesora de la misma.

Al Consejo Nacional de Ciencia y Tecnología por el apoyo recibido mediante la beca para realizar estudios en el programa de maestría en didáctica de las Matemáticas.

REFERENCIAS

[1] C. Díaz & I. de la Fuente, I. "Razonamiento sobre probabilidad condicional e implicaciones para la enseñanza de la estadística". *Épsilon*, vol. 59, pp. 245-260. 2005

[2] G. Mejía, L. Y. Sierra, F. Fernández. "Influencia del contexto y la estructura en la actuación de los estudiantes al resolver problemas de probabilidad condicional," *Revista de Didáctica de las matemáticas, vol.* 86, pp. 95-100. 2014.

[3] C. Batanero, J. D. Godino, D. R. Green, P. Holmes & A. Vallecillos. "Errors and difficulties in understanding elementary statistical concepts", *International Journal of Mathematics*, vol. 25, No. 4, pp. 527-247. 1998.

[4] E. Sánchez. "La probabilidad en el programa de estudio de matemáticas de la secundaria en México". *Educación matemática*, vol. 21, No. 2, pp. 39-77. 2009.

[5] J. E. Tarr y J. K. Lannin. "How can teachers build notions of conditional probability and independence?," *Exploring probability in school: Challenges for teaching and learning*". New York: Springer. (pp.216-238). 2005.

[6] C. Díaz & I. de la Fuente. "Validación de un cuestionario de razonamiento probabilístico condicional". *Revista electrónica de metodología aplicada*, vol. 12 No. 1, pp.1-15. 2007.

[7] Y. Chevallard. "El análisis de las prácticas docentes en la teoría antropológica de lo didáctico". *Researches en Didactique des Mathématiques*, vol. 19, No. 2, pp. 221-266. 1999.

[8] O. León, I. Montero. "Métodos de investigación en psicología y educación", Mc Graw Hill. España, 2003.

[9] F. Hitt, A.S. Gonzalez-Martín. "Covariation between variables in a Modelling process: the ACODESA (collaborative learning, scientific debate and self-reflection) method". *Educational Studies in Mathematics*, No. 88, pp. 201-219. 2015.

[10] A.M. Ojeda. "Dificultades del alumnado respecto a la probabilidad condicional". *UNO*, vol. 5, pp. 37-55. 1995

[11] J. I. Barragués & J. Guisasola. "Una propuesta para la enseñanza de la probabilidad en la universidad basada en la investigación didáctica". Educación matemática, col. 21 No. 3, pp. 127-162. 2009

[12] M. A. Lonjedo & M. P. Huerta. "La resolución de problemas de probabilidad condicional: un estudio exploratorio con estudiantes de bachiller". Departamento de Didáctica de la Matemática, Universidad de Valencia (Memoria de Tercer Ciclo no publicada)

SUGEY TATIANA SOTOMAYOR

UNIVERSIDAD AUTÓNOMA DE QUERÉTARO

DOCENTESUGEYSOTOMAYOR@GMAIL.COM

02

UNA PROPUESTA DE ACTIVIDADES PARA LA COMPRENSIÓN DE LA DERIVADA EN LOS PROBLEMAS DE OPTIMIZACIÓN

A PROPOSAL FOR ACTIVITIES TO ACHIEVE THE UNDERSTANDING OF THE DERIVATIVE IN OPTIMIZATION PROBLEMS

RESUMEN

El presente trabajo pretende que, a través de los problemas de optimización, el estudiante pueda acercarse a la comprensión de la noción de derivada. Para tal fin, se propone una actividad que consiste en que el alumno resuelva un problema de optimización desde diferentes registros de representación y justifique cada uno de sus procedimientos, lo cual le permitirá evidenciar la implicación de la derivada en la resolución del problema planteado. El uso de los distintos registros de representación para la resolución del problema planteado corresponde con las diversas maneras de realizar una tarea, las cuales están definidas como técnicas en una praxeología matemática. De esta forma, el análisis a priori del presente trabajo se hace considerando la Teoría de Registros de Representación Semiótica y la Teoría Antropológica de la Didáctica.

Palabras Clave: 000, Teoría Antropológico de Lo Didáctico, problema de optimización, derivada, registros de representación., actividad.

ABSTRACT

This paper intends that, through optimization problems, the student will be able to reach an understanding of the notion of derivative. For this purpose, an activity is proposed, which consists in the student solving an optimization problem from different representation registers and justifying each one of his or her procedures, this will allow him or her to show the implication of the derivative in the resolution of the proposed problem. The use of the different representation registers for the solution of the proposed problem agrees with the different ways of solving an assignment, which are defined as techniques in a mathematical praxeology. In this way, the a priori analysis of this work is done considering the Theory of Registers of Semiotic Representation and the Anthropological Theory of Didactics.

Keywords: 000, Anthropological Theory of the Didactic, optimization problem, derivative, semiotic representation registers, activity.

INTRODUCCIÓN

La mayoría de los estudiantes resuelven ejercicios donde se les pide calcular la derivada de una determinada función real, casi siempre de forma automática, aplicando las reglas usuales de derivación sin realmente concientizarse de las implicaciones de este concepto matemático. Por tanto, se puede decir que una de las razones por las cuales ocurre esta sistematización continua, se debe a que prevalece la visión sistemática algebraica de la derivada, tanto en el cálculo de funciones en escenarios descontextualizados, como en escenarios de problemas de optimización.

Con respecto a estos últimos, a los estudiantes se les dificulta traducirlos del lenguaje natural al algebraico, lo cual les impide realizar la formulación adecuada de un modelo matemático, que les permita encontrar una solución óptima al problema. Cabe mencionar que la mayoría de los docentes enseñan este contenido matemático haciendo uso solamente de procesos algorítmicos y algebraicos, y no utilizan otro tipo de representaciones. Por tal motivo, los alumnos resuelven los problemas de optimización utilizando los criterios de máximos y mínimos de una función (representación algebraica), sin analizar ni justificar sus procedimientos.

Dentro de las investigaciones centradas en los problemas de optimización, se destaca el estudio realizado por Baccelli, Anchorena, Figueroa y Prieto [1], quienes mencionan que las dificultades presentadas por los estudiantes están asociadas al planteamiento, análisis y resolución de estos problemas. Estos obstáculos se evidenciaron en el análisis que realizaron al momento de observar la manera en cómo estudiantes de primer año de la Facultad de Ingeniería de la Universidad Nacional de Mar del Plata (Argentina) resolvían problemas de este tipo.

En este mismo orden de ideas, se presenta la investigación realizada por Aldana [2], quien menciona que el conocimiento alcanzado por los estudiantes sobre el concepto de optimización es de tipo intuitivo, ya que está asociado a la gráfica de la función; es decir, que entienden la optimización como el punto más bajo o más alto de la gráfica, para lo cual aplican el criterio de la segunda derivada para encontrar los máximos o mínimos.

Por otra parte, Rojas, Báez y Corona [3] presentan una propuesta didáctica para la enseñanza del tema de optimización, apoyada con Excel y Geogebra para estudiantes de bachillerato. Los autores consideran que, si desde el inicio del curso de cálculo diferencial se planteara un problema contextualizado sobre optimización, el avance cotidiano de conocimientos capacitaría al estudiante para entenderlo y resolverlo.

En este contexto, Navarro [4] propone una secuencia didáctica para la construcción del concepto de derivada en problemas de optimización; la actividad se desarrolló con estudiantes que ya habían tomado un curso de cálculo diferencial y, a través del uso de un manipulable físico y uno digital como el software de Geogebra, los estudiantes lograron una noción intuitiva de la derivada de una función al tener una aplicación en problemas de optimización de contexto extra-matemático que resolvieron sin conocer la algoritmia ni la palabra "derivada".

Por otro lado, Cruzado [5] presenta una investigación que tiene como finalidad analizar de qué manera estudiantes de las diferentes carreras de ingeniería coordinan registros de representación semiótica al resolver problemas de optimización movilizando el concepto de derivada de funciones reales de variable real. Este estudio toma como marco teórico la Teoría de Registros de Representación Semiótica. Para la parte experimental de la investigación se elaboraron dos problemas de optimización mediados por el Geogebra, los cuales fueron aplicados a estudiantes de Ingeniería Mecánica de una universidad nacional peruana. El análisis de los resultados logrados por los estudiantes evidenció que hay dificultades al momento de coordinar el registro figural, algebraico, gráfico y en lengua natural. Se concluyó que los problemas de optimización propuestos favorecen para que se dé dicha coordinación.

Las investigaciones previas pretenden poner de manifiesto las dificultades que encuentran los estudiantes al resolver problemas de optimización. Pero, sobre todo, la persistencia de éstas debido a la implicación del concepto de derivada. De esta manera, con la intención de que los estudiantes alcancen una comprensión de la derivada, se hace el presente trabajo, el cual consiste en el diseño de una secuencia de actividades cuya finalidad es que los alumnos puedan establecer una conexión entre los diferentes registros de representación.

MARCO TEÓRICO

La presente propuesta se fundamenta en dos marcos teóricos: la Teoría de Registros de Representación Semiótica (RRS) y la Teoría Antropológica de la Didáctica (TAD).

Teoría de Registros de Representación Semiótica

Según Duval [6], aprender matemáticas involucra una secuencia de actividades cognitivas como la conceptualización, el razonamiento y la resolución de problemas; es por ello que, para aprender y enseñar matemáticas, se requiere la utilización de distintos registros de representación y de expresión: el lenguaje natural, la representación mediante imágenes, símbolos, etc. Por tal motivo, la enseñanza-aprendizaje del concepto de derivada por medio de los problemas de optimización no se debe trabajar solo haciendo uso de un único registro, sino que se debe incluir la capacidad de traducir la información de una representación a otra.

De igual forma, el autor afirma que los objetos matemáticos no son accesibles a la percepción o la experiencia intuitiva inmediata, ya que no son objetos reales o concretos, por tal razón, se hace necesario disponer de las diferentes representaciones semióticas de un objeto; sin embargo, no se debe confundir el objeto con su representación, distinción que es fundamental para su comprensión.

Con base en lo anterior, Duval [6] indica que un registro es un campo de variación de representación semiótica en función de los factores cognitivos que le son propios. Es necesario señalar que las representaciones semióticas son producciones que hace un sujeto a partir de sus representaciones mentales para representar un objeto matemático mediante signos que tienen su propia significación y funcionamiento dentro de un sistema de representación. En otras palabras, las representaciones se pueden considerar como un medio para exteriorizar las concepciones mentales con la intención que se dé una comunicación con otros sujetos.

En la teoría de registros de representación semiótica, a la actividad ligada a la producción de una representación se le llama semiosis, mientras que a la aprehensión conceptual de los objetos matemáticos se denota como noesis. Un registro de representación debe permitir las tres actividades cognitivas ligadas a la semiosis: la formación de una representación identificable, el tratamiento y la conversión. La primera de estas se refiere a la expresión mental, es decir a la expresión de un objeto en un determinado registro semiótico, lo cual implica que se debe seleccionar un conjunto de caracteres, además de las relaciones y datos que permiten constituir lo que representamos. La siguiente es la transformación, la cual se refiere a una transformación interna, es decir es la transformación de la representación en el mismo registro en el que está dada. La tercera y última es la conversión, la cual hace alusión a una transformación externa, o sea, la representación en un registro distinto al registro en el que fue dada. Por tanto, se puede decir que estas dos últimas actividades están relacionadas con la transformación de las representaciones en otras representaciones Duval [6].

La Teoría Antropológica de la Didáctica

La Teoría Antropológica de la Didáctica (TAD) de Chevallard [8] parte del principio que el saber matemático se construye como respuesta al estudio de cuestiones problemáticas, apareciendo así como el resultado de un proceso de estudio. Dicho proceso, en cuanto actividad que conduce a la construcción de conocimiento matemático, forma parte de la actividad matemática.

La TAD identifica lo didáctico con todo lo relativo al estudio, tomando la palabra “estudio” en un sentido muy amplio, que engloba las nociones de enseñanza y aprendizaje comúnmente utilizadas en la cultura pedagógica, y se refiere a todo aquello que se hace en una determinada institución para aportar respuestas a las cuestiones o para llevar a cabo las tareas problemáticas que se plantean.

Dentro del punto de vista general del conocimiento matemático, se propone la noción de organización praxeológica matemática o praxeología matemática (o simplemente organización matemática) como modelo básico para describir el conocimiento matemático. Dicho saber está organizado en dos niveles:

- El primer nivel es el que remite a la práctica que se realiza, la praxis o saber-hacer, es decir, a los tipos de problemas o tareas que se estudian y las técnicas que se construyen y utilizan para abordarlos.
- El segundo nivel recoge la parte descriptiva, organizadora y justificadora de la actividad, que llamaremos logos o simplemente saber. Incluye las descripciones y explicaciones que se elaboran para hacer inteligibles las técnicas, esto es, el discurso tecnológico (la razón, logos, de la técnica y, en última instancia, el fundamento de la producción de nuevas técnicas) y la teoría que da sentido a los problemas planteados; este nivel permite interpretar las técnicas y fundamentar las descripciones o demostraciones tecnológicas.

Para fines de este trabajo nos centraremos en el primer nivel de la praxeología matemática y en las tecnologías. Los tipos de tareas son: interpretar, representar, trazar, completar y determinar; las técnicas están relacionadas con el uso de los diferentes registros de representación mientras el alumno resuelve el problema de optimización; y en cuanto a las tecnologías, de acuerdo a la actividad propuesta, se irán presentando o construyendo sobre las tareas.

A continuación, se presenta un esquema que muestra la interacción de los registros de Representación y las praxeologías matemáticas en la resolución de un problema de optimización.

Tabla 1. Relación de los Registros y las Praxeologías en las actividades

Registros	Tipo de tarea	Técnica	Tecnología
Lengua natural	T_1: Interpretar	t_1: leer el problema propuesto e interpretarlo.	
Figural	T_2: Representar	t_2: Dibujar un rectángulo dividido por la mitad. t_3: Señalar las dimensiones del rectángulo en el dibujo.	
Algebraico	T_1:Interpretar T_4:Determinar	t_4: Reconocer cómo se calcula el perímetro y el área de un rectángulo. t_5: Establecer la relación de la cantidad de malla necesaria, la cual es el perímetro del rectángulo y la división del centro. Es decir, 2 bases y 3 alturas $corral = 2b + 3h$. t_6: Expresar una relación matemática con la información del problema, la cual es el perímetro de 300m, $300 = 2b + 3h$. t_7: Despejar cualquiera de las dos variables de la ecuación anterior y sustituir dicha variable en la ecuación del área del corral. t_8: Derivar la función del área e igualar a cero para determinar el valor máximo. t_9: Sustituir este valor en una de las variables que se despejó, ya sea base o altura. t_{10}: Encontrar la otra medida. t_{11}: Determinar las dimensiones del terreno.	θ_1: $perímetro = 2(a + b)$ siendo a y b, los dos lados diferentes del rectángulo θ_2 : $Área\ del\ rectángulo = bh$ θ_3 ::Si n es un entero positivo y $f(x) = xn$, entonces $f'(x) = nxn^{-1}$ θ_4 :Si $F(x) = f(x) - g(x)$ entonces$F'(x) = f'(x) - g'(x)$ θ_5 : sean f y f' derivables en a. Entonces a es máximo relativo de f si: $f(a) = 0$ y $f''(a) < 0$
Gráfico	T_2: Representar T_1: Interpretar T_4: Determinar T_5:Trazar	t_{12}: Construir una tabla de valores de x y y. t_{13}: Representar gráficamente los valores de la tabla en un plano cartesiano. t_{14}: Realizar una interpretación de la gráfica. t_{15}: Determinar cuál es el intervalo en el que x puede tomar valores para que el área alcance el valor más alto. t_{16}:ubicar el punto máximo en la gráfica. t_{17}: Trazar rectas tangentes a la gráfica de la función, dos antes del valor máximo encontrado y dos después del mismo. t_{18}: Trazar una recta tangente en el valor máximo hallado.	θ_6 : Una recta L que pase por P $(a, f(a))$, se denomina recta tangente a la gráfica de f en P, si L es la mejor aproximación lineal de f cerca de P. θ_7 : La derivada en un punto x=c es la pendiente de la recta tangente a la gráfica de f en el punto (c, f (c)). θ_8 : Sea $f: X \rightarrow Y$, con X, $Y \subset$ R. El conjunto $G(f)= (x, y): x \in X, y = f(x) \subset R \times R = R2$ se llama gráfica de la función .
Numérico	T_6: Completar T_1: Interpretar T_4: Determinar	t_{19}: Completar la tabla propuesta. t_{20}: Realizar una interpretación de la tabla, una vez haya sido completada. t_{21}: Determinar los valores de las dimensiones, para los cuáles el área es máxima.	

Fuente: Elaboración propia

METODOLOGÍA

En este apartado se muestra el diseño de la actividad, la cual consiste en la presentación de un problema de optimización, luego, a través de una serie de indicaciones se orienta al estudiante hacia la resolución del problema.

Problema: Un ranchero tiene 300 m de malla para cercar dos corrales rectangulares iguales y contiguos; es decir, comparten un lado de la recta. Determinar las dimensiones de los corrales para que el área cercada sea máxima.

Sección 1: En esta sección se plantea la necesidad de realizar una interpretación del problema. De esta manera, al estudiante se le solicita lo siguiente: Realice una representación geométrica del problema de acuerdo con las condiciones dadas.

Esta sección tiene como finalidad que el estudiante pase del registro natural al registro figural, es decir, el estudiante debe hacer una transición entre dichos registros y, para esto, se espera que el alumno interprete geométricamente el problema planteado. El alumno deberá realizar una representación como la figura1. Asimismo, en términos de la TAD se puede decir que esta sección promueve la técnica geométrica.

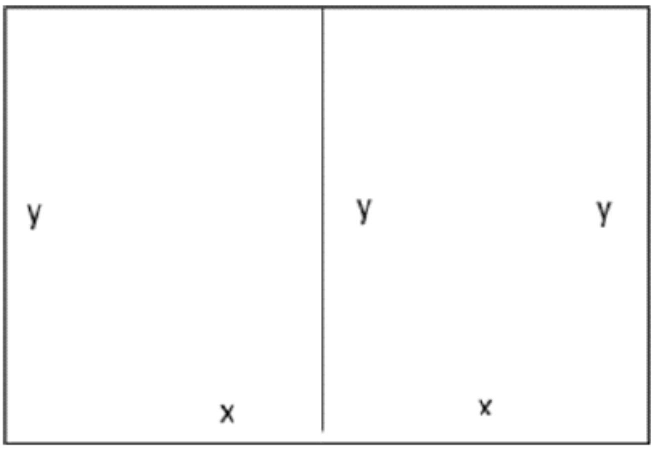


Figura 1. Representación figural del problema propuesto
Fuente: Elaboración propia

Sección 2: En esta sección se plantea la necesidad de una interpretación algebraica. De este modo, se solicita al estudiante lo siguiente: Determine la expresión matemática que modela el área cercada por los dos corrales en función de una sola variable.

La finalidad de esta sección es que el estudiante haga la transición del registro figural al algebraico. Para ello, al igual que en la sección anterior, el estudiante debe hacer una adecuada interpretación geométrica del problema planteado, ya que sin dicha interpretación le será difícil determinar la expresión matemática que modela el área de los corrales, la cual sería la representación algebraica del problema de optimización. Esta sección promueve la técnica algebraica.

Una posible representación algebraica sería:

$$Área = f(x) = 2x\left(\frac{300-4x}{3}\right) = 200x - \frac{8}{3}x^2 \qquad (1)$$

Sección 3: En esta sección se plantea la necesidad de una interpretación gráfica. De este modo, se le solicita al estudiante lo siguiente: Represente gráficamente la expresión definida para la Ec. (1) y diga cuál es el intervalo en el que x puede tomar valores para que el área alcance los valores más altos. Justifique su respuesta.

El objetivo de esta sección es que el estudiante pase del registro algebraico al registro gráfico y después al registro natural. Para ello, el estudiante debe tener claro qué significa la coordenada de un punto en la gráfica de la Ec (1). Si el estudiante responde que la ordenada del punto representa el área de los corrales para determinado valor de la dimensión del corral, entonces se puede decir que el estudiante realiza la transición entre dichos registros. Esta sección promueve la técnica gráfica.

Sección 4: En esta sección se plantea la necesidad de una interpretación numérica. De este modo, se le pide al estudiante lo siguiente: Complete la siguiente tabla y determine los valores de las dimensiones de los corrales para que el área cercada sea máxima. Argumente su respuesta.

Tabla 2. Área de los corrales en función de una sola variable

x	$y = (300-4x)/3$	$A = f(x) = 2x\left(\frac{300-4x}{3}\right)$
	100	0
10		
15		2400
	60	
	0	0

La finalidad de esta sección es que el estudiante transite del registro algebraico al registro numérico y luego al registro natural. Para tal objetivo el estudiante debe determinar los valores de las dimensiones de los corrales para que el área cercada sea máxima. Esta sección promueve la técnica numérica.

Sección 5: En esta sección se plantea la necesidad de una interpretación gráfica. De este modo, se le solicita al estudiante lo siguiente: Ubique en la gráfica el valor máximo del área.

La finalidad de esta sección es que el alumno coordine el registro numérico con el registro gráfico. Para tal fin, el estudiante debe hacer una interpretación adecuada de la tabla 1, al igual que de la gráfica de la Ec (1). Si el estudiante es capaz de determinar el valor máximo del área, se puede decir que comprende el significado del valor óptimo pedido. Esta sección promueve la técnica gráfica.

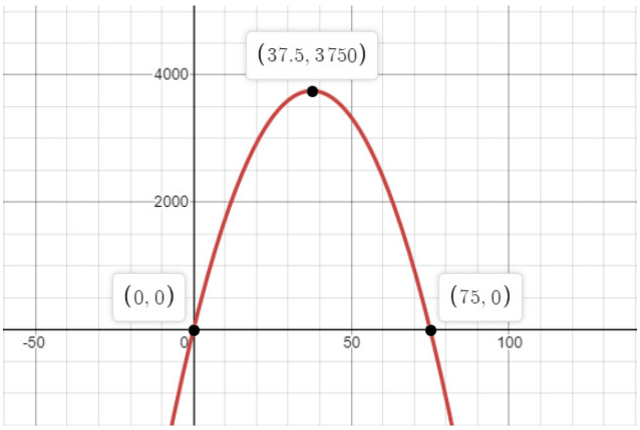


Figura 2. Representación gráfica del problema propuesto
Fuente: Elaboración propia

Sección 6: Esta sección plantea la necesidad de una interpretación grafica de la función. De esta manera, se le solicita al estudiante lo siguiente: Trace tangentes a la gráfica, dos antes del valor máximo encontrado y dos después del mismo. ¿Qué puedes decir acerca de sus grados de inclinación? Explique.

Esta sección tiene como finalidad que el estudiante pase del registro gráfico al registro natural. Si el estudiante proporciona un argumento válido con relación a las tangentes trazadas a la función del área, la transición entre dichos registros será la adecuada. Esta sección promueve la técnica natural.

Sección 7: Esta sección plantea la necesidad de una interpretación gráfica. De este modo, se le propone al estudiante que conteste a una serie de cuestiones asociadas a la gráfica.

a) Trace una tangente a la gráfica en el valor máximo hallado. ¿Qué puedes decir acerca de su grado de inclinación?

b) Explique qué relación tienen las rectas tangentes trazadas con el valor óptimo del área.

c) Derive la función hallada en la sección 2 y diga para qué valores se anula. Compare el valor con el del hallado en la sección 5. Justifique su respuesta.

d) Evalúe la derivada en un punto antes del hallado anteriormente y en uno después ¿Qué puede decir del signo de las derivadas?

e) Finalmente, explique qué relación tienen las tangentes a la gráfica de la función con la derivada de la misma.

Al igual que en las secciones anteriores, esta última sección tiene como finalidad que los estudiantes transiten en los registros de representación y utilicen diversas técnicas para resolverlos; pero a diferencia de las otras, le permitirá al estudiante acercarse a una noción más amplia de la derivada en comparación con la que emplea usualmente, la cual es vista solo desde un registro algebraico.

RESULTADOS Y DISCUSIÓN

Con base en las secciones que orientan al estudiante para resolver el problema, se espera que los alumnos encuentren dificultad al momento de determinar la expresión matemática que modela el área de los corrales en función de una sola variable, esto se debe a que, en muchas ocasiones, al estudiante se le facilita la función a optimizar y no se trabaja para que sean ellos los que la construyan. Además, puede presentarse el caso de que no tenga claro el significado de "perímetro de la superficie", y se equivoque al momento de expresarlo en función de las dos dimensiones.

Por otro lado, si el alumno no realiza una interpretación adecuada de la gráfica que representa el área de los corrales, puede errar al momento de completar la tabla que corresponde a la sección 4 y, por consiguiente, en la argumentación de su respuesta.

Finalmente, se espera que el estudiante evidencie que la noción que tiene de la derivada está asociada a un registro algebraico y no a otro. Con relación a esto, es probable que muestre falencia al contestar preguntas de la sección 7.

CONCLUSIONES

La enseñanza del Cálculo centrado en el dominio de algoritmia genera en los estudiantes una pobre comprensión del significado de los conceptos y de su aplicación al momento de resolver un problema que no se les presente sintetizado algebraicamente. En el caso que nos ocupa, el estudiante está acostumbrado a que el docente le proporcione ejercicios o problemas sobre derivadas, los cuales, probablemente resolverán de manera mecánica, guiándose de los ejemplos explicados en clases y en los que solo cambiarán algunos datos, pero la forma de darle solución no será distinta de la que ya conocen.

Respecto a esta problemática, se hizo el presente trabajo para promover la comprensión de la derivada a través de los problemas de optimización. La manera como está estructurada la actividad le permitirá al estudiante transitar de un registro de representación a otro, y de esta forma visualizar distintas formas de realizar el problema de optimización planteado, logrando así una comprensión más profunda de la noción de la derivada; es decir. el estudiante no solo podrá contemplarla en su registro algebraico, sino que además podrá comprender su interpretación geométrica y, de esta manera, visualizar la implicación que tiene en la resolución de problemas de optimización.

AGRADECIMIENTOS

Se agradece al Consejo Nacional de Ciencia y Tecnología por el apoyo recibido mediante la beca para realizar estudios en el programa de maestría en didáctica de las Matemáticas.

REFERENCIAS

[1] S. Baccelli, S. Anchorena, S. Figueroa, and G. Prieto, "Problemas de optimización: un análisis en la construcción de significados" in *Congreso Iberoamericano de Ciencia, Tecnología, Innovación y Educación*, 2014, p. 12.

[2] L. Gallego and E. Aldana, "Análisis de la concepción de la actividad de optimizar, desde una ingeniería didáctica" *Educ. científica y técnológica*, vol 216 pp. 225–228, 2013.

[3] L. Rojas, J. Báez, and M. Corona, "Propuesta didáctica para la enseñanza del tema de optimización, apoyado con Excel y Geogebra, para estudiantes de bachillerato" *El Cálculo y su Enseñanza, Enseñanza las Ciencias y la Matemática*, vol. 9, pp. 52–63, 2017.

[4] L. Navarro, "Secuencia didáctica para la construcción del concepto derivada en problemas de optimización" Instituto Tecnológico de Sonora, 2014.

[5] E. Cruzado, "Problemas de optimización mediados por el Geogebra que movilizan el concepto de derivada de funciones reales de variable real en estudiantes de ingeniería" Pontificia Universidad Católica del Perú, 2018.

[6] R. Duval, "*Semiosis y pensamiento humano*" (M. Vega, Trad.) Universidad del Valle, Instituto de Educación y Pedagogía, Grupo de Educación Matemática. Cali, Colombia 2004.

[7] Y. Chevallard, "El análisis de las prácticas docentes en la teoría antropológica de lo didáctico" *Rec. en Didact. des Mathéematicques*, vol. 19, no. 2, pp. 221–266, 1999.

ILIANA MARÍA PATERNINA ORTEGA
EDUARDO CASTAÑO TOSTADO
MARIO SANTANA CIBRIÁN

UNIVERSIDAD AUTÓNOMA DE QUERÉTARO

PATERNINAO@GMAIL.COM

03

UN ANÁLISIS COMPARATIVO DE MODELAJE ESTADÍSTICO EN ESTUDIOS DE VIDA DE ANAQUEL SENSORIAL

A COMPARATIVE ANALYSIS OF STATISTICAL MODELING IN SENSORY SHELF LIFE STUDIES

RESUMEN

El presente artículo describe un modelo estadístico para datos provenientes de análisis sensoriales de alimentos, teniendo en cuenta la estructura de agrupación de tipo clúster que éstos poseen, así como el patrón de asociación en el tiempo para el conjunto de datos. Se hace uso del enfoque de las ecuaciones estimantes generalizadas, que ajusta modelos de media marginal con la ventaja de que solamente es necesaria la especificación correcta de las medias marginales para que los estimadores de los parámetros sean consistentes y asintóticamente normales. Para una base de datos específica donde se evalúa sensorialmente un yogurt, se encontró que la probabilidad de rechazo esperada sobre el tiempo es no lineal y creciente; reportándose un 13 % de rechazo poblacional esperado a las cero horas de almacenamiento del producto.

Palabras Clave: IMA, vida de anaquel sensorial, clúster, autocorrelación, ecuaciones estimantes generalizadas, marginal.

ABSTRACT

This article describes a statistical model for data recovered from sensorial food analyses, taking into account the cluster-type grouping structure that they have, as well as the pattern of association over time for the data set, adjusting a marginal model using the generalized estimating equations approach, a method that adjusts marginal mean models with the advantage that only the correct specification of the marginal means is necessary for the parameter estimators to be consistent and asymptotically normal. For a specific database, where a yogurt is sensorially evaluated, it was found that the probability of rejection over time is nonlinear and increasing; reporting a 13% expected population rejection at zero hours of product storage.

Keywords: IMA, sensory shelf life, cluster, autocorrelation, generalized estimating equations, marginal.

INTRODUCCIÓN

El análisis de sobrevivencia es el conjunto de herramientas estadístico-matemáticas que se han utilizado en las Ciencias de los Alimentos para realizar las inferencias respectivas a la vida de anaquel y la vida de anaquel sensorial de alimentos [1], [2], [3] y [4]. Hough *et al.* [1] fueron los primeros en aplicar análisis de sobrevivencia para estimar la vida de anaquel sensorial considerando la naturaleza censurada de los datos; de allí en adelante, se ha aplicado tal metodología, lo que ha resultado en la creación de diversas aplicaciones [2] y [3].

En estudios de vida de anaquel sensorial, la evaluación hecha por evaluadores humanos para rechazar o aceptar un producto alimenticio en diferentes tiempos de almacenamiento juega un papel especial. Las respuestas de cada evaluador al paso del tiempo de almacenamiento del producto alimenticio son registradas y, aunque espaciadas en el tiempo, pueden considerarse como un clúster de datos (datos longitudinales), además de que éstos pueden contener un patrón de asociación en el tiempo, de manera que, en este tipo de estudios y en una primera aproximación, será pertinente considerar que los datos de cada evaluador forman un clúster con una estructura de autocorrelación y pueden ser usados en el modelaje estadístico.

Por otro lado, cabe señalar que, aparte de la estructura de agrupación y de autocorrelación mencionadas, existe la posibilidad de considerar otras circunstancias en el modelaje que representan factores inherentes en los individuos, para reconocer así diferencias entre evaluadores; por ejemplo, características sociodemográficas que pueden influenciar la evaluación del producto. Aquellos factores y otros inherentes de los individuos serán considerados en futuras investigaciones.

Para incorporar los anteriores aspectos y analizar datos provenientes de análisis sensoriales de alimentos, se hará uso del análisis de datos longitudinales, lo que permitirá modelar la probabilidad de rechazo por parte de los consumidores a través del tiempo. Generando así alternativas de modelaje estadístico para estudios de vida de anaquel sensorial.

Así las cosas, en el presente trabajo se lleva a cabo un modelaje estadístico de datos provenientes de análisis sensorial de alimentos, incorporando la estructura de agrupación de tipo clúster que éstos poseen, así como el patrón de asociación en el tiempo para el conjunto de datos.

FUNDAMENTOS TEÓRICOS

La característica que define un estudio longitudinal es que los individuos son medidos repetidamente a través del tiempo, en contraste con los estudios transversales, en que una sola respuesta es medida para cada individuo. Los estudios longitudinales pueden distinguir cambios en el tiempo dentro de los individuos (*ageing effects*) de las diferencias entre las personas en sus líneas base (*cohort effects*).

Los datos longitudinales pueden ser recolectados prospectivamente, siguiendo los sujetos a través del tiempo; o retrospectivamente, extrayendo múltiples medidas de cada persona de un archivo histórico. En el caso de los estudios sensoriales de alimentos, éstos se pueden recolectar de manera prospectiva, que corresponde a un diseño tipo básico, y realizando las diferentes evaluaciones en un solo momento del tiempo, conocido como diseño en reversa, ver [5] para mayores referencias.

Hay tres enfoques para el modelaje de datos longitudinales discretos y continuos: usando extensiones de los modelos lineales generalizados, que son modelos marginales, modelos de efectos aleatorios y modelos de transición. En el presente trabajo se usa el enfoque de los modelos marginales.

En los modelos marginales, la regresión de la variable respuesta, Yij, sobre las variables explicativas es modelada separadamente de la correlación dentro de las personas, dado que los valores repetidos probablemente no son independientes; este análisis debe incluir también hipótesis acerca de la forma de la correlación. El enfoque de modelo marginal tiene la ventaja de modelar separadamente la media y la covarianza. En la regresión, modelamos la esperanza marginal, (Yij), en función de las variables explicativas sin tener en cuenta la dependencia entre observaciones. Este tipo de modelaje es bastante práctico y llamativo para un científico de alimentos porque le dará información concerniente del efecto del tiempo de almacenamiento sobre la aceptabilidad o rechazo del producto.

En esta primera aproximación del uso de datos longitudinales en estudios de vida de anaquel sensorial, tiene sentido físico suponer que las respuestas repetidas de cada consumidor pueden no tener una dependencia fuerte entre ellas, debido a que los científicos de alimentos diseñan sus pruebas sensoriales con el fin de que la respuesta de un consumidor en una i-ésima cata, no se vea alterada por las anteriores o posteriores, ejemplo de ello es la utilización de diseños en reversa, la utilización de cabinas sensoriales especiales para las degustaciones y el uso de limpiadores de paladar, entre otros [5].

Específicamente, un modelo marginal tiene los siguientes supuestos:

a. La esperanza marginal de la respuesta, $(Yij)=\mu ij$, depende de las variables explicativas $\mathbf{xij}$ por h(μij) = $\mathbf{xij}'\boldsymbol{\beta}$, donde h es una función de enlace conocida, tal como logit para las respuestas binarias o log para conteos;

b. La varianza marginal depende de la media marginal de acuerdo a $Va(Yij)=\upsilon(\mu ij)\phi$, donde υ es una función de varianza conocida y ϕ es un parámetro de escala que podría ser estimado.

c. La correlación entre Yij y Yik es una función de la media marginal y quizás de parámetros adicionales $\boldsymbol{\alpha}$, es decir, $Cor(Yij,Yik) = \rho(\mu ij,\mu ik;\boldsymbol{\alpha})$, donde $\rho(\cdot)$ es una función conocida.

De manera que el modelo marginal está dado por:

$$logit(\mu_{ij}) = log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right) = log\left(\frac{P(Y_{ij}=1|T=t_{ij})}{1-P(Y_{ij}=1|T=t_{ij})}\right) = \alpha + \beta * t_{ij},\ i = 1,2,\dots,n\ y\ j = 1,2,\dots,m. \quad (1)$$

donde $Yij = 1$ denota que el i-ésimo consumidor rechazó la muestra del producto en el j-ésimo tiempo de almacenamiento, tij corresponde al j-ésimo tiempo de almacenamiento del i-ésimo consumidor, α y β son los parámetros a estimar del modelo que son calculados a través de las ecuaciones estimantes generalizadas, GEE por sus siglas en inglés, Generalized Estimating Equations.

Las GEE, son un método general para analizar datos recolectados en clústeres donde:

• Las observaciones dentro de un clúster pueden estar correlacionadas.

• Las observaciones en clústeres separados son independientes.

• Una transformación monótona de la esperanza está linealmente relacionada a las variables explicativas.

• La varianza es una función de la esperanza.

Como se mencionó antes, hay varios enfoques para modelar datos longitudinales, y así extender los modelos lineales generalizados (GLM, por sus siglas en inglés, Generalized Lineal Models) a datos longitudinales, los modelos de efectos mixtos y de transición que especifican completamente la distribución conjunta dentro de los clústeres vía variables latentes o dinámicas condicionales; pero con la presencia de efectos aleatorios, la estimación de verosimilitud necesita la integración sobre las distribuciones de efectos aleatorios, que quizás sean numéricamente intratables. Así que, comparados con estos enfoques, el método GEE ajusta modelos de media marginal, con la ventaja de que solamente es necesaria la especificación correcta de las medias marginales para que los estimadores de los parámetros $\hat{\alpha}$ y $\hat{\beta}$ sean consistentes y asintóticamente normales.

En la metodología de las ecuaciones estimantes generalizadas, el usuario puede fijar una estructura de correlación, que con frecuencia es llamada, matriz de correlación de trabajo. Algunas estructuras de correlación de trabajo incluyen, intercambiable, la independiente, no estructurada y autoregresiva.

METODOLOGÍA

Para los análisis estadísticos, se hace uso del paquete geepack [5] disponible en el *Software* R; todas las tablas y figuras presentadas son de creación propia.

El paquete *geepack* trabaja bajo el enfoque de las Ecuaciones Estimantes Generalizadas, mencionadas antes (GEE por sus siglas en inglés, Generalized Estimating Equations), que son un método general para analizar datos recolectados en clústeres que surgen de medir repetidamente a los individuos a través del tiempo. A continuación se describe el uso de dicha función colocando los nombres en inglés, ya que así está predeterminado en el *software* R.

LA FUNCIÓN GEEGLM

La función principal que utiliza la biblioteca *geepack* de R para hacer las estimaciones de los parámetros correspondientes al modelo estadístico, es la función *geeglm*, que entre otros argumentos tiene los siguientes:

family: la función de varianza es especificada por el argumento *family* y es identificada por el nombre de la distribución correspondiente en un modelo lineal generalizado. En la Tabla 1 se especifican las familias más representativas con las que se cuentan y sus respectivas funciones de varianza, (μ).

Tabla.1 Opciones del argumento *family* en la función *geeglm*.

Nombre	Función de varianza
Gaussian	Identity
Binomial	$\mu(1-\mu)$, $\mu \in (0, 1)$
Poisson	μ, $\mu > 0$
Gamma	$\mu 2$, $\mu > 0$

corstr: las estructuras de correlación de trabajo ("*working*") predefinidas son especificadas con este argumento, en la Tabla 2 se muestran las diferentes opciones de estructuras de correlación y sus respectivas funciones de correlación, $R(\alpha)$.

Tabla 2. Opciones del argumento *corstr* en la función *geeglm*.

Nombre	Función de correlación
Independence	COR$(Yit,Yit') = 0$, $t \neq t'$
Exchangeable	COR$(Yit,Yit') = \alpha$, $t \neq t'$
ar1	COR$(Yit,Yit') = \alpha\|t-t'\|$, $t \neq t'$
Unstructured	COR$(Yit,Yit') = \alpha tt'$, $t \neq t'$

Los datos que provienen de estudios sensoriales de alimentos y a los cuales se les realizan los análisis estadísticos en el presente trabajo son de dominio público y se pueden encontrar en [5], a su vez, se pueden descargar en archivo .xlsx de la página web del editor.

Para la generación de la base de datos, los tecnólogos de alimentos básicamente seleccionan una muestra de consumidores, a los cuales se les pide prueben un conjunto de muestras de cierto alimento, con diferentes tiempos de almacenamiento, y respondan "sí" o "no" a la pregunta "¿normalmente consumiría este producto?"; de esto, generan un conjunto de datos como los mostrados en la Tabla 3.

Tabla 3. Ilustración de un conjunto de datos que se generan en análisis sensoriales de alimentos

	Tiempos de almacenamiento					
Consumidor	t_0	t_1	t_2	t_3	…	t_n
1	no	no	si	si	…	no
2	si	si	si	si	…	no
3	si	no	si	no	…	si
4	no	si	si	no	…	no
5	si	si	si	si	…	no
⋮	⋮	⋮	⋮	⋮	⋮	⋮
N	si	No	no	no	…	no

RESULTADOS Y DISCUSIÓN

Un modelo marginal logístico con la misma estructura media y la función de varianza binomial $(\mu it) = \mu it(1 - \mu it)$, utilizando el enfoque GEE, es ajustado con diferentes estructuras de correlación; los resultados obtenidos se muestran a continuación:

Utilizando una estructura de correlación independiente:

	intercepto	tiempo
Estimación	-1.86816	0.07457
Error estándar	0.20197	0.00856
Wald	85.6	76.0
Pr(>|W|)	<2e-16 ***	<2e-16 ***

Utilizando una correlación intercambiable:

	intercepto	tiempo
Estimación	-1.8431	0.0745
Error estándar	0.2020	0.0086
Wald	83.2	75.0
Pr(>|W|)	<2E-16 ***	<2E-16 ***

Utilizando una correlación autorregresiva de primer orden:

	intercepto	tiempo
Estimación	-1.88795	0.07497
Error estándar	0.20124	0.00868
Wald	88.0	74.7
Pr(>|W|)	<2E-16 ***	<2E-16 ***

Utilizando una correlación no estructurada:

	intercepto	tiempo
Estimación	-1.92043	0.07236
Error estándar	0.19851	0.00804
Wald	93.6	80.9
Pr(>|W|)	<2E-16 ***	<2E-16 ***

A su vez, se ajusta un modelo logístico mediante el enfoque de modelos lineales generalizados, lo que sería apropiado si no hubiera en los datos la estructura de tipo clúster, y no hubiera dispersión en las probabilidades de respuesta para los consumidores con los mismos valores de covariables; los resultados se muestran abajo:

	intercepto	Time
Estimación	-1.86816	0.07457
Error estándar	0.21080	0.00845
Wald	--8.86	8.83
Pr(>|W|)	<2E-16 ***	<2E-16 ***

Todos los resultados anteriores se resumen en las Tablas 4 y 5, y se puede apreciar que, en todos los casos, las estimaciones de los parámetros de cada modelo ajustado son cercanas y los errores estándar son aproximadamente iguales.

Para seleccionar el mejor modelo, Pan propuso un método de selección de modelo para GEE, que llamó Criterio de Información de Cuasiverosimilitud (QIC por sus siglas en inglés Quasilikelihood information criterion) [7] y [8], que también puede ser utilizado para seleccionar la mejor estructura de correlación de trabajo en el análisis GEE. Los resultados del coeficiente QIC se resumen en la Tabla 6.

Tabla 4. Estimaciones de los parámetros de los modelos ajustados.

	intercepto	time
Estimación	-1.87*	0.07*
Error estándar	0.21	0.01
Estimación	-1.87[1]	0.07[1]
Error estándar	0.20	0.01
Estimación	-1.84[2]	0.07[2]
Error estándar	0.20	0.01

*Usando el enfoque de modelos lineales generalizados y el enfoque de las ecuaciones estimantes generalizadas con estructura de correlación: 1Independiente, 2Intercambiable.

Tabla 5. Estimaciones de los parámetros de los modelos ajustados.

	intercepto	Time
Estimación	1.89[3]	0.07[3]
Error estándar	0.20	0.01
Estimación	-1.92[4]	0.07[4]
Error estándar	0.20	0.01

Usando el enfoque de las ecuaciones estimantes generalizadas con estructura de correlación: 3Autorregresivo de orden uno y 4No estructurado.

Tabla 6. Criterio de información de cuasiverosimilitud para los diferentes modelos ajustados.

Modelo	Estructura de correlación	Criterio de información de cuasiverosimilitud
modelo.uns	no estructurada	375
modelo.ar	Autorregresivo de orden 1	376
modelo.ind	independiente	376
modelo.exc	intercambiable	376

La estructura de correlación no estructurada tuvo el valor más pequeño de QIC, aunque no difiere en gran medida de los demás, se puede considerar que el modelo bajo la estructura de correlación no estructurada es el mejor. Además, en general, sí el número de unidades por clúster es pequeño en un diseño balanceado y completo, entonces una matriz no estructurada es la recomendada [7] y [8].

Teniendo en cuenta lo anterior, escogemos el modelo con estructura de correlación no estructurada, y así un modelo marginal análogo a la Ec. (1), vendría dado por:

$$logit(\hat{\mu}_{ij}) = log\left(\frac{\hat{\mu}_{ij}}{1-\hat{\mu}_{ij}}\right) = -1.92 + 0,07t_{ij}, i = 1,2,\dots,50 \; y \; j = 1,2,3,\dots,7. \qquad (2)$$

donde Yij=1 denota que el i-èsimo consumidor rechazó la muestra del producto en el j-èsimo tiempo de almacenamiento. De manera que:

$$\hat{\mu}_{ij} = \hat{P}(\boldsymbol{Y} = 1|\boldsymbol{T} = t) = \frac{e^{-1.92+0,07*t}}{1+e^{-1.92+0,07*t}} \qquad (3)$$

Por tanto:

$$\hat{P}(\boldsymbol{Y} = 1|\boldsymbol{T} = t) = \frac{e^{-1.92+0,07*t}}{1+e^{-1.92+0,07*t}} \qquad (4)$$

El modelo estimado en la expresión (4) se puede representar mediante la Fig. 1, que representa la probabilidad de rechazo estimada en el tiempo.

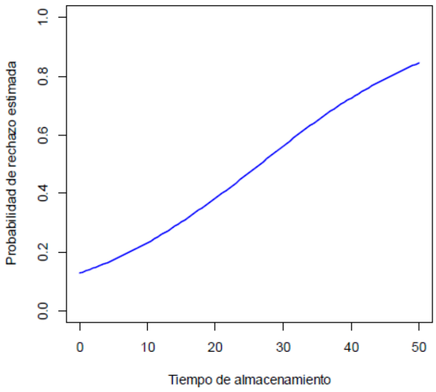


Figura 1. Estimación de la probabilidad de rechazo por parte de los consumidores en el tiempo.

Con el enfoque GEE se puede verificar que, en el tiempo cero, ya hay una probabilidad de rechazo estimada aproximado del 13%, que nos lleva a pensar que el modelo marginal reconoce la historia completa del consumidor respecto a su percepción del producto a través del tiempo, a diferencia del modelo para datos censurados, que otorga un 0% de rechazo al tiempo cero de almacenamiento, esto se puede ver en la Fig 2.

En particular, al tiempo cero existen cuatro consumidores que manifestaron rechazo, consumidores que son retirados del análisis estadístico realizado con el enfoque de análisis de sobrevivencia; en cambio, en el modelaje aquí presentado no fue necesario retirarlos, de manera que el modelo marginal tiene en cuenta la posibilidad de que haya consumidores que no acepten el producto fresco pero que puedan aceptarlo eventualmente.

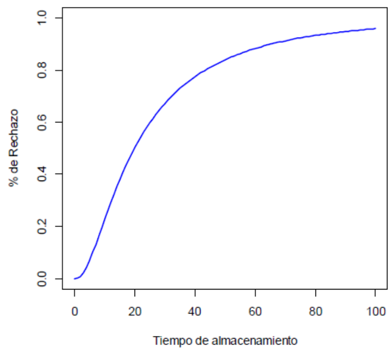


Figura 2. Función de distribución acumulada para la vida de anaquel sensorial.

La Tabla 7 muestran las estimaciones de las vidas de anaquel estimadas mediante el modelaje de datos censurados para 10, 20 y 25 % de rechazo.

Tabla 7. Estimaciones de la vida de anaquel sensorial y sus intervalos de confianza del 95%.

vida de anaquel sensorial (horas)	Error estándar	Límite Inferior	Límite Superior	Porcentaje de rechazo
6.03	1.33	3.91	9.30	10
9.08	1.67	6.33	13.02	20
10.60	1.83	7.56	14.87	25

Mediante el análisis de datos longitudinales, se estiman las probabilidades de rechazo para las vidas de anaquel sensorial de 6, 9 y 11 horas, que se pueden ver en la Tabla 8.

Tabla 8. Estimaciones de la probabilidad de rechazo y sus intervalos de confianza del 95%.

Vida de anaquel sensorial (horas)	Estimación probabilidad de rechazo	Error estándar	Límite Inferior	Límite Superior
6	0.18	0.027	0.13	0.24
9	0.22	0.030	0.16	0.28
11	0.25	0.031	0.18	0.31

Claramente, los resultados encontrados con el enfoque de análisis de datos longitudinales, difieren de los encontrados con el análisis de sobrevivencia, debido a que se modelan variables aleatorias completamente diferentes. Con el modelo aquí presentado, podemos responder a la pregunta de cuál es la probabilidad de rechazo poblacional cuando el tiempo de vida de anaquel sensorial fuera especulado por el analista de alimentos.

Sin embargo, se puede notar que los resultados generados por el modelo marginal reflejan que, por debajo de las 8 horas de almacenamiento, las probabilidades de rechazo son superiores a 0.15, lo que representa una mayor probabilidad de rechazo en las primeras horas de almacenamiento que el modelo de análisis de sobrevivencia; por otro lado, para tiempos de almacenamiento por encima de las 12 horas, las probabilidades de rechazo son menores comparadas con los porcentajes de rechazo reportados por el modelaje de datos censurados.

CONCLUSIONES

El modelo marginal permite modelar las respuestas directas de los consumidores, evitando la censura de las mismas y presuponiendo que ello generará estimaciones más realistas de la percepción sensorial que tiene el consumidor en torno al producto alimenticio.

AGRADECIMIENTOS

Los autores agradecen al Consejo Nacional de Ciencia y Tecnología por su apoyo durante toda esta investigación.

REFERENCIAS

[1] G. Hough, K. Langhor, G. Gòmez, and A. Curia, "Survival analysis applied to sensory shelf-life of foods", *J. Food Sci*. Vol. 68, Nr. 1, pp. 359-362, 2003.

[2] D.A. Jacobo-Velasquez, P.A. Ramos-Parra, and C. HERNANDEZ-BRENES, "Survival analysis applied to the sensory shelf-life dating of high hydrostatic pressure processed avocado and mango pulps", *J. Food Sci*. Vol. 75, Nr. 6, pp. 286-291, 2010.

[3] A. Cruz, E. Waltert, R. Silva, J. Faria, H. Bolini, H. Pinheiro, and A. Santana, "Survival analysis methodology to predict the shelf-life of probiotic flavored yogurt", *Food Res*. Vol. 43, Issue 5, pp. 1444-1448, 2010.

[4] A. Giménez, F. Ares, G. Ares, "Sensory shelf-life estimation: A review of current methodological approaches", *Food Research International*. Vol. 49, Issue 5, pp. 311-325, 2012.

[5] G. Hough, *Sensory Shelf Life Estimation of Food Products*. UK: Taylor and Francis Group. 2010.

[6] U. Halekoh, S. Højsgaard, J. Yan, "The R Package geepack for Generalized estimating Equations", *Journal of Statistical Software*, Volume 15, Issue 2., pp. 1-11, 2006.

[7] W. Pan, "Akaike's Information Criterion in Generalized Estimating Equations", *Biometrics* Vol. 57, pp. 120-125, 2001.

[8] W. Pan, "Goodness of fit Tests for GEE with Correlated Binary Data", Scand J Statist Vol. 29, pp. 101-110, 2002.

[9] Y. Kwon, T. Park, A. Ziegler, and M. Paik, "Generalized estimating equations with stabilized working correlation structure", Computational Statistics and Data Analysis, Vol. 106, pp.1–11, 2017.

[10] Z. Liu, Methods and Applications of Longitudinal Data Analysis. Elsevier Inc. 2016.

La presente edición de
PädiUAQ
fue maquetada por Alejandro Zamorano
en la Coordinación de Diseño e Imagen de la Facultad de Ingeniería
de la Universidad Autónoma de Querétaro.
El cuidado estuvo a cargo de Daniela Pérez y Soid Ruiz
Se publicó en junio de 2020.
en Santiago de Querétaro, México.